BUTTERWORTH HEINEMANN

Utilities Policy, Vol. 5, No. 3/4, pp. 219–236, 1995

0957–1787/96 $15.00 + 0.00

S0957–1787(96)00005–7

Market power in California electricity markets

Severin Borenstein, James Bushnell, Edward Kahn and Steven Stoft

As the electricity industry in California undergoes a process of fundamental restructuring, important new products and markets will be created while others will lose significance. In this paper, we undertake an initial survey of the products and markets that will be prominent in the emerging new electricity industry. We describe approaches to analyzing the prospects for, and the impacts of, market power abuse in these various product markets. The key product markets that are discussed include those for spot electrical energy, for pool-based and physical power contracts, and for reliability services such as load balancing and spinning reserve. Structural measures of market power, such as the Hirschman–Herfindahl Index (HHI), have certain general shortcomings that are exacerbated when applied to restructured electricity markets. Fortunately, the direct estimation of competitive equilibria, such as a Cournot oligopoly equilibrium, appears to be more feasible for this industry than is generally the case.

The US electricity industry is undergoing a process of fundamental change. If current regulatory timetables are adhered to, a dramatically redesigned market for electric energy will be operating in California within the next 2 years. This introduction of a new, pool-based, spot market for electricity will coincide with the deregulation of California electricity suppliers and the unbundling of the set of products traditionally known as 'electric service.' While many key details of California's emerging markets have yet to be finalized, it is clear that in this process new products and markets will be created while others will lose significance or disappear completely. One of the many challenges currently confronting policy makers is to determine the competitive outlook for these markets and to identify those products that may be vulnerable to the abuse of market power in the restructured industry. To address this issue, we conducted an initial survey of the product markets that will be central to the new emerging electricity industry. For each of these markets, we describe approaches for analyzing the prospects for, and the impacts of, market power abuse.

Severin Borenstein, James Bushnell, Edward Kahn and Steven Stoft are with the University of California Energy Institute, 2539 Channing Way, Berkeley CA 94720, USA.

Product markets – an overview

Since so many details of the coming California electricity marketplace are still unknown, we must conduct our survey in the context of a reasonable, but stylized, representation of the future electricity marketplace. In this section, we detail some of the key aspects of our stylized representation.

The central feature of our stylized market is a power exchange. The exchange will accommodate suppliers and consumers in an auction in which bids take the form of supply and demand curves, indicating quantities for a spectrum of prices. The exchange will determine market clearing prices and quantities for several key locations on the grid, based upon these bids and the physical characteristics of the network. The effects of transmission congestion and losses will be made manifest in the locational price differences between these nodes on the grid. We assume that the customers of the exchange will initially be distribution companies formed from the current major utilities of the state and some number of 'large' customers who are allowed access directly to the exchange and find it economical to do so. We also assume that the distribution companies (Distcos) will retain some affiliation with the generation companies (Gencos) formed from the assets of the utilities.

We assume that an Independent System Operator (ISO), existing in parallel with the exchange, will ensure grid stability and enforce the network feasibility of transactions that are reached through the exchange or other market-making processes such as direct contracts for 'physical' delivery of energy. Our analysis in this paper does not require an explicit model of the process through which physical contracts and the exchange are reconciled with network constraints. The potential impact of such a reconciliation process on market power issues is discussed below.

The task of defining the product markets stemming from a future California electricity exchange is complicated by its lack of predecessors in the USA.[1] The existence of an exchange alters the electricity marketplace so drastically that it makes moot some historically important electricity product distinctions while also creating new ones. However, despite the lack of domestic examples to fall back upon, defining products in electricity is greatly simplified by the fact that the existence of a power exchange continues the evolution of 'electricity' into that of a pure commodity. At its fundamental level, electric energy itself is a non-differentiated product, but it is still differentiated by time and location.

Both demand and production cost vary over time. While there is a continuum of demand states that arise over time, we simplify the analysis by considering only two states, which we call *peak* and *off-peak*. The distinction made between the two states is based on the possibility that the geographic market for electrical energy may be different for these two cases due to transmission congestion. A more detailed analysis would most likely require further differentiation and demand levels, depending upon the effects of congestion and the time-varying composition of demand.

The creation of a spot electricity market will create markets for new products that were previously bundled with electricity sales. One new market that would be created by the existence of an exchange is the market for contracts. This market would include exchange-based financial instruments such as contracts for differences as well as direct contracts for the physical delivery of power. Transmission congestion contracts are a specialized pool-based financial instrument that, although they fall under the general paradigm of a contract market, may require individual attention with respect to market power issues. The competition conditions of the contract market will have a direct influence on competition in the underlying electricity spot market, particularly on the ability of new players to enter those markets. The unbundling of electric service will also necessitate the creation of several markets that we will lump together under the general classification of ancillary grid and reliability services. Such services would include spinning reserve, load balancing, and voltage support.

Table 1.

	Product class	
Spot energy	**Contracts**	**Reliability services**
Off-peak electricity	Contracts for differences	Spinning reserve
On-peak electricity	Physical power contracts	Load balancing
	Transmission congestion contracts	Voltage support

These are the product markets that seem to be most heavily influenced by, and of the most concern to, policy makers, and to which we will devote the bulk of this paper. These products are summarized in Table 1.

Reliability service versus capacity payments. The nature of product markets depends upon the overall market structure: what may be a relevant product in an industry under one market structure can become irrelevant under another market structure. This distinction applies to what has been called the capacity (or 'firm power') market under the traditional bulk power market structure as we shift to a pure spot market. Joskow[2], for example, makes this point in his comments on the FERC Mega NOPR. In a pure spot market, no one assures services. Hence there is no obligation to supply, nor is there any corresponding financial compensation, or capacity payment, for this service.

It has been pointed out, however, that electricity differs from most other goods, because the product is not easily storable, and from most other services, because states of excess demand for even brief periods of time can be extremely costly. Thus, there is a need for short-notice supply increases or demand decreases in the event of unexpected shortfalls. In this paper, we call these *reliability services*. Suppliers who chose to provide these services would be responsible for acquiring the capacity to do so. Demanders who offered to be curtailed on short notice would be responsible for their own costs and obligations if such a contingency arose and their service was curtailed. The product in question is therefore reliability services, not capacity *per se*. Investment in the production of these services could therefore be driven, as it is in most other industries, by the expected revenues from providing these services.

It has been argued[3] that generation capacity in an electricity system constitutes a public good, and that some form of capacity payment is therefore necessary to overcome this market failure. However, this argument relies upon the assumption that supply will continue to be randomly rationed during periods of shortfall. If demand is responsive to the price increases that result from generation shortfalls, then only the loads with the lowest value for power (values that are less than the current cost of production) will not be served. This eliminates the externality created by excess capacity. Implementing such responsiveness would be a significant change from traditional procedures which have relied upon random rationing.

Since the spot market structure that is most familiar in electricity is the one in the UK, and that market includes a capacity payment, it is worth examining that market briefly to try to understand what problem the designers of the market thought they were addressing, and how the market has performed.

The capacity payment in the UK is based on the formula VOLL multiplied by LOLP, where VOLL is the value of lost load (estimated at 2/kWh in 1990 and escalated with consumer prices thereafter) and LOLP is the loss of load probability. The intention behind the capacity payment was to provide a signal for new investment beyond the spot market price. In practice, the capacity payment has proven to be quite volatile. This should have been expected, since LOLP is a very sensitive function that is nearly exponential in available capacity.[4]

A volatile capacity payment may provide extra revenue for new entrants, but it does not address the risk issues that are fundamental to new investment. There are two ways to finance new investment: either the new entrant uses the strength of its corporate borrowing power (ie its balance sheet) to borrow money for new facilities, or it borrows on the strength of a long-term contract with customers. Balance sheet finance is quite limited compared with financing based on long-term contracts; ie there are far fewer entities capable of the balance sheet approach. If there were no long-term contract market, there would be potential market power problems due to entry barriers. The capacity payment mechanism in the UK did not contribute directly to new entry; that has happened primarily through the contract market. However, it is difficult to determine the extent to which the capacity payment has influenced contract price.

The capacity payment in the UK has proven to be quite vulnerable to manipulation, however, due primarily to its specific dependence on LOLP. Since LOLP can change quite rapidly, it creates exactly the kind of price structure in which withdrawal of capacity can be profitable. Since the UK system provides little or no demand-side incentives for price response, the LOLP increases very rapidly when the load nears system capacity. In those cases, the incentive for a firm to withdraw some generation from the market is large as long as it still has enough generation active to gain from the price increase. There were several incidents involving manipulation of the capacity payment in 1991 that were discussed by the Office of Electricity Regulation (OFFER). More recently, the December 1994–January 1995 winter peak saw extremely high capacity payments amounting to nearly US$1.5 billion, or about 20% of wholesale revenue. These were due to unplanned outages at two nuclear stations and one of National Power's coal stations. As a result, the average pool price for the 1994–1995 year will increase over the previous year, despite the efforts of OFFER to cap the energy price from the spot market. These facts are summarized in Newbery[5], where a model of the incentive to manipulate the capacity payment is also developed. This model shows that with only modest market shares (on the order of 10%), withdrawal of capacity can be profitable.[6]

If there is no explicit payment for capacity, will the market supply the optimal level of reliability? The real-time nature of decision-making and response necessity leads some observers to believe that the market will be less than likely to attain the efficient level of reliability. While that certainly would be true if all negotiations and bidding had to be done in real time when a shortfall occured, it does not follow once we recognize that firms can negotiate contingent contracts in advance of such occurrences. On the supply side, this is the role of spinning reserves. A buyer, distributor, system operator, or generator could have already contracted for the spinning reserves before they become necessary. The spinning reserve bid or contract could be non-linear, involving a payment for availability of the service and an additional payment when the service is actually used.[7] On the demand side, interruptible users provide the equivalent of short-run price sensitivity that is unlikely to occur in real time. Different types of consumers will be willing to bear different levels of potential interruption in exchange for either fixed or marginal payments, but in any case, consumers will be able to 'contract' their price elasticity in advance. Such real-time response does not require real-time negotiations. It does however, require real-time communication and control. These systems are already in place on the supply side, and their adaptation to the demand side will be crucial to the development of efficient, decentralized generation investment.

Unresolved issues about future electricity markets

Many important aspects of the future electricity market are not included in our stylized representation. Some of these details will have important effects on the competitive nature of the various electricity product markets. A more complete analysis of market power issues will require either further information or the making of assumptions about these details.

A key factor in determining the behavior of Distcos in the newly formed markets is the regulatory mechanisms used to determine the revenues of these companies. Of particular import is whether Distcos will pass through to their customers the cost of their spot market purchases or whether some form of price cap incentives are used. The same questions apply to the incentives of the Distcos in the contract markets. Will Distcos be allowed to profit from signing advantageous CFDs? Or will these gains also be passed through to ratepayers?

Similar incentive questions arise about the regulatory mechanisms used to recover the sunk costs of utility stranded assets. The extent and manner of cost recovery implemented in the restructured industry will have an important impact on the competitive position of the utility-affiliated Gencos relative to other independent suppliers. For the purposes of market power analysis, it is unclear how the independently owned generating units currently under contract to the utilities should be treated. Should this generation be viewed as potential competition to the large Gencos? Or should it be treated as part of the Gencos' supply portfolios? The answer will depend in a large part on how the existing contracts are adapted to the restructured market.

The role of physical power contracts. Under the California Public Utilities Commission's 20 December 1995 decision,[8] physical energy trades would be allowed to exist in parallel with the exchange-based spot market. Since the details of this decision are still unresolved, it is difficult to determine whether or not a distinction should be made, for market power analysis purposes, between this 'direct-access' market and the exchange market.

The market power implications of a physical contract market can be viewed on two levels. The first assumes that the power exchange and physical contract markets have no significant structural inefficiencies outside of those related to market power. Under this assumption the two product markets should not be considered distinct for the purposes of market power analysis. These two products would be near-perfect substitutes for each other, so that artificially high prices in one market would involve buyer substitution to the other. Conversely, the existence of market power in one market would imply that it exists in the other. A monopoly supplier, for example, will still be able to exercise market power whether it sells directly to buyers or through a power exchange.

The bulk of this paper operates on this first level of analysis. Most of our analysis does not depend upon the energy spot markets being organized in the exchange-based form we have described. An analysis of the potential for market power abuse in the supply of the central product in this industry, electric energy, will have to examine the costs and capacities of the major electricity generators in the market as well as estimates of demand elasticities for this product. At this most basic level of analysis, the exact nature of the market-clearing process is not modeled, and is in fact implicitly assumed to be efficient. Similarly, physical limitations on markets, such as transmission congestion, are independent of the specifics of the market process. That is to say, besides the possible exercise of market power, market inefficiencies such as information asymmetries, irrational behavior, or sub-optimal dispatch algorithms are not considered. In this way the debate over the relative efficiencies of various market processes is separated from the analysis of the potential for the abuse of market power. Past experiments with competition, such as the auction process for independent power supply contracts in California, have shown that a poorly implemented market process can be subject to inefficiencies even with a high level of competition,[9] while economic theory tells us that even an efficient market process can be vulnerable to market power abuse.

The second, more detailed, level of analysis concerns the institutional barriers that affect the interplay between the exchange-based spot market and the physical power contract market. It is at this deeper level of analysis that the potential impact of physical power contracts will need to be considered. Market power analysis at this level would be concerned with more than just the costs and capacities of generators and elasticities of consumers. Factors affecting the ability of customers to transact, such as transmission policies that favor either physical contracts or exchange transactions, would also have to be considered. These are also important issues, but they cannot be addressed until more detail about the coordination of bilateral contracts with the exchange-based markets is known.

Naturally, an analysis of the competitive prospects for some of the other key products that we describe, such as reliability services, will depend upon the details of the market process. The very existence of a market for some of these products will depend upon the details of restructuring. In these areas, we are forced to rely more upon our forecasts of how these markets will evolve.

Organization of this paper

The remainder of this paper is organized according to the product markets we analyze. First we discuss general approaches to analyzing market power concerns, and establish a framework for our analysis. Next we examine the market for off-peak spot electricity, and discuss several models of oligopoly behavior that could be applied to this market. We then address the issue of transmission congestion, which may force geographic distinctions in the markets for on-peak spot electricity, and discuss the market for physical power contracts and exchange-based financial instruments. Finally, we analyze the prospects for competition in the various emerging markets for reliability services such as voltage support.

Market power analysis

The most obvious concern about market power in a restructured electricity market is the possibility that one or a group of suppliers of electricity could restrict output and raise the wholesale price. This is a potential concern in both the spot electricity markets, peak and off-peak, and the markets for reliability services. As we discuss below, the ability of one or a collection of firms to exercise market power depends not just on concentration, but also on the ease with which new firms can enter, and smaller incumbent firms can expand their output in the market. Even if a small set of firms supplies the majority of output, if other firms can easily increase their quantity in the market, then the dominant firm(s) will have little ability to raise price. Thus, the measurement of the market power of sellers should begin with some measure of concentration, but it must proceed well beyond that point.

While the structure of the generation market after restructuring is a central area of concern, an analysis of market power must also take into account the transmission and distribution sectors. Since Distcos will be the principal buyers, and will be local monopolists for some time, the way in which these companies are regulated will greatly affect the incentives of both buyers and sellers of power. If a Distco can automatically adjust retail prices to pass through any wholesale (exchange) price fluctuations, then there is little incentive for it to take actions that could lower wholesale electricity prices. If the Distco were under a price cap without an adjustment (or a complete adjustment) for wholesale prices, however, the firm could have a strong incentive to encourage lower wholesale prices.

The analysis will be further complicated if the Distcos retain some ownership of generating facilities. This changes a Distco's incentives, particularly if it is allowed to pass through changes in the wholesale price. A Distco that owns significant generation might be able to increase its profits at the wholesale level while losing very little at the retail level by restricting its wholesale output and driving up the price.

Some concern also has been expressed that the vertical integration of Distcos could give them an incentive to use their distribution divisions to harm competitors in the generation market. As discussed shortly, the Distco is a natural supplier of financial instruments that could lower the risk of building new generation. The availability of such hedging instruments could facilitate new entry into the generation market, entry of firms that would compete directly with the Distco's affiliated generation. Thus, Distcos could have an incentive to reduce the availability of low-price 'insurance' from hedging instruments, thereby increasing the cost of financing for potential new entrants.

Finally, transmission will play a critical role in the analysis of market power. Simply taken as an exogenous factor in the competitive environment, transmission capacity will determine the boundaries of electricity markets. If there is sufficient transmission capacity between two locations, then they will behave as one market. If not, then the ability of a producer in one market to participate in the other will be limited. Furthermore, most transmission capacity will continue to be owned by companies that also generate electricity and distribute the product to consumers. If firms engaged in generation and distribution are also the major source of investment for new transmission facilities, these firms may have incentives to manipulate such investment to benefit their other divisions at the expense of competing sellers and buyers.[10]

As this overview of market power indicates, the issue is quite complex in the electricity industry. Geographic market definitions will depend on transmission constraints, which will vary with load, and may be determined by firms that also own generation and distribution assets. Within a market, firms will have differing incentives to try to raise or lower the wholesale price, which will depend on the degree of vertical integration and the ability to hedge price risk in the market. Though standard measures of concentration provide some information about the potential for market power abuse, it is clear that they cannot capture some of the most important information necessary for the analysis.

Even for the generation market, standard structural measures suffer two serious shortcomings. First, traditional 'market share' measures, based on historical sales, are of questionable value since the nature of the market after deregulation will be so radically changed. Second, other structural measures that don't rely on historic sales, such as generating capacity, do not account for the relationship between capacity and demand or the relative cost curves of competitors. These features would, to some extent, be reflected in measures based upon historical sales, if those were relevant for the restructured industry. In particular, capacity-based measures do not incorporate the extent to which independent generating capacity and imports can meet demand and whether the marginal cost of that capacity is competitive with that of dominant firms.

The most widely used measure of concentration in a market is the Hirschman–Herfindahl Index (HHI), which is defined as the sum of the squared market shares. One appeal of the HHI is that it is linked directly to market power in one theoretical model of competition.[11] Two factors, in general, determine the level of market power that a firm can exercise: the elasticity of demand in a market and the degree of competition among sellers. In perfect competition, the elasticity of demand becomes irrelevant due to the intensity of competition. In monopoly, only the demand elasticity matters since there are no competitors. In an oligopoly setting, the actions of competitors and elasticity of demand both influence the market outcome. An important measure of market outcomes is the price–cost margin. In one of the standard models of oligopoly, symmetric Cournot competition, the price–cost margin will be:

$$\frac{P - MC}{P} = \frac{1}{n\epsilon} = \frac{\mathrm{HHI}}{\epsilon} \tag{1}$$

where P is price, MC is marginal cost, ϵ is the elasticity of market demand, and n is the number of identical firms. Thus, the HHI measures directly one of the two factors that determine the exercise of market power, but it gives no indication of the elasticity of demand and, therefore, a very imperfect indication of the severity of the market power problem. In this case, the HHI indicates by how much price exceeds marginal cost relative to the outcome that would result under monopoly (HHI = 1). The HHI also implicitly assumes that all competitors are behaving in similar, oligopolistic, fashion. As is demonstrated in the following section, this assumption can be misleading when the market contains many small 'fringe' suppliers whose aggregate capacity is significant. Predicting oligopolistic equilibria can be difficult, and often requires a great deal of proprietary data, while computing an HHI is often fairly straightforward. The HHI therefore can be a more practical way of establishing the same conclusions predicted by an equilibrium analysis. Unfortunately, the HHI can also mislead if the world does not conform to the simple assumptions of the symmetric Cournot model.

Equation (1) also highlights the importance of price-responsive demand. If consumers cannot get the information or technology to respond to short-run price changes, then the effective demand elasticity will be very low. With a very low elasticity, prices can exceed cost substantially, even in relatively unconcentrated markets. For instance, with a demand elasticity of 0.2, Equation (1) indicates that 10 equal-sized firms will result in a price 100% above marginal cost. Alternatively, in order to fall within the Department of Justice Guidelines for competition – price not exceeding the competitive level (marginal cost) by more than 5% – this market would require about 100 equal-sized firms.

Problems with structural measures: an example

To illustrate the potential problems with the HHI and other structural measures we turn to a simple model of a dominant firm facing many smaller competitors. We begin by making two simplifying assumptions. In modeling a market for spot electricity in a single period, we assume that demand for that period is both completely inelastic and is known with certainty. In the short run prices actually will be implemented by matching supply bids to (roughly) day-ahead forecasted load. Interruptible loads linked to spot prices would provide price sensitivity that would produce a downward-sloping market demand curve.

Despite the inelastic market demand in the short run, we assume that the dominant firm faces a price-sensitive firm-level demand, because of the presence of a competitive 'fringe' of sellers. Each fringe supplier is not large enough to influence price on its own, but taken as an aggregate, they can have a significant effect on the behavior of the dominant firm. In the standard model, the dominant firm selects a production quantity and assumes that the

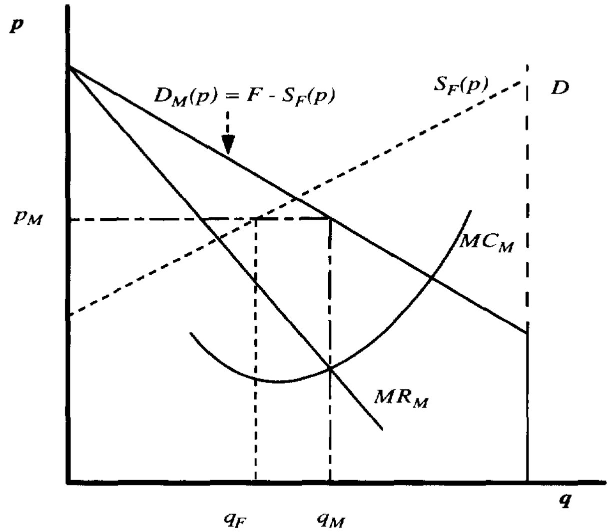

Figure 1. Monopoly supplier facing a competitive fringe

fringe will produce at a level that equates the fringe's marginal cost to the resultant market clearing price. When demand is totally inelastic (see the vertical line labeled D in Figure 1) and the fringe supply curve, S_F, is upward sloping, this has the effect of creating a demand curve faced by the dominant firm, $D_M(p)$, that is downward sloping. The dominant firm then sets quantity according to the standard monopoly solution by setting its production level, q_M, at the point where its marginal cost, MC_M, intersects the marginal revenue curve, MR_M, that is associated with the 'residual' demand it faces. The final market clearing price, p_M, therefore depends upon the marginal cost curves of both the dominant firm and the competitive fringe.

For instance, consider the outcome in such a market if instead of being upward sloping, the fringe supply curve were very low out to q_F and then vertical at q_F. In that case, the dominant firm's optimal price would be infinite and its output would be q_M. Its market share, and any measure of market structure, would be the same as in the case illustrated in Figure 1, yet the price and margin would be infinitely higher.

This is an extreme example, though a similar result will occur if the fringe supply curve is simply very steep. The market place will be much higher than in Figure 1 even though the market shares will be the same. This illustrates how the ability of large firms to manipulate prices depends upon more than just the firms' relative sizes. The cost curves and capacity of the competitive fringe will affect the slope and intersection of the residual demand curve perceived by the large firm. If that residual demand is very elastic, the large firm's ability to affect prices will be limited.

In reality, even a dominant firm would be constrained not only by the presence of a competitive fringe, but also by the longer-run elasticity of demand. If a dominant firm raised prices significantly, consumers would respond in the medium and long run by reducing their energy use, relying on alternative energy sources, or installing the metering and control equipment necessary to make their demand price-sensitive. It is worth noting that even a firm that faces very inelastic demand in the short run will still take into account the longer-run effects of price gouging when it sets its price. The degree to which concern about long-run response will prevent a dominant firm (or group of large firms) from exploiting very inelastic short-run demand is very difficult to estimate.

Analysis of California electricity markets

In the context of emerging electricity markets the data to be used in structural measures such as an HHI are either unavailable because the markets have not yet emerged or inapplicable because deregulation will have changed the incentives of key players. Fortunately, much more cost and demand data are available for these markets than for most others. Thus, while structural indexes appear to be much harder to evaluate in this case, estimates of competitive equilibria appear to be more feasible for this industry than is generally the case.

Of course, assembling the data necessary for estimates of oligopoly equilibria is not without its difficulties. The detailed cost data that would be required for estimating the cost curves of future competitors have been supplied by the existing utilities for various regulatory proceedings. Estimating the costs and capacities of competitors outside of California will be somewhat more challenging. While the rated 'capacities' of transmission lines that carry bulk power into the state are known, the true impact capabilities of these lines in fact depend on both transmission schedules and the load conditions of the entire Western Systems Coordinating Council. These capabilities can therefore vary widely. Seasonal and weather conditions will also affect the competitive impact of imported power.

Demand data present the greatest challenge to estimating competitive equilibria in electricity. As we have discussed, very short-run elasticities are currently quite low. Without the development and implementation of technologies that can make demand more price-responsive, this is likely to lead to indications of significant market power. Thus, it will be important to analyze the potential market penetration of these technologies and the degree to which firms are likely to take longer-run response into account.

Given the problems with using structural measures in this industry and the availability of detailed cost data, we start out by exploring the theoretical models that might inform a market power analysis, including the Cournot model. To complete the full analysis, one must identify the nature of competition and the incentives of the sellers and buyers in the market. In the following sections, we outline the analysis of the key product markets, the nature of competition for those markets, and some of the key factors that will influence competition. Where appropriate, we discuss theoretical models from which empirical analyses can be developed.

The market for off-peak spot power: a unified market

Standard market-power analysis assumes the existence of a single unified market in which all participants trade at a single price. This description is a useful approximation of the problem at hand under two circumstances. First, it is useful under low-demand conditions when the entire California market is essentially free of congestion and thus can be treated as a single node. Second, it is useful when a region becomes insulated from the competitive forces in other regions by strongly congested lines. In this case one must, of course, take into account the effective shift in demand caused by the power inflow on the congested lines. This section develops the theory of market power for the electricity industry in a unified (one-node) market, possibly with imports, while the following section addresses problems that arise when weak or congested lines are present within the market under analysis.

A primary use of the methods developed in this section will be to analyze the market for off-peak power in the restructured industry. This market will include at least the four major generators in the state, PG&E, SCE, SDG&E, and LADWP, as well as the potential for imports from other WSCC states. It is unclear how the export capacities of neighboring states should be measured in relation to the native load requirements of those states. At the very least, these imports create a significant competitive fringe in the California market. These considerations suggest that in the off-peak, market power will be severely limited by competitive forces, and indicates the tools that should be used to confirm that conjecture.

Oligopoly with a competitive fringe

In a market in which a small number of companies are responsible for a significant share of sale, the perfect competition model of firms as price takers is not relevant. In such markets, the relevant models of competitive behavior are oligopoly models. We begin by discussing the standards: the Cournot and Bertrand models. In both models, competitors reach a Nash equilibrium in which each firm optimizes based on the assumption that the other will continue with its chosen strategy. In the Cournot model, firms compete by setting sales levels and then accepting whatever market price results. In the Bertrand model, they compete by setting prices and then supplying whatever quantity is demanded of them. With homogeneous goods, Cournot competition can produce prices significantly above competitive levels, depending on the elasticity of demand as discussed in the previous section, while the outcome of Bertrand competition is the competitive market price.

Because the models result in very different outcomes, any attempt to analyze an oligopoly market must confront the question of which model is more applicable. The answer depends on the market. In some markets, firms can set a price and then produce (or withdraw from inventory) whatever quantity is demanded from it. In other markets, firms must schedule production and then accept whatever price allows it to sell its output.

In the context of an electricity exchange, the Cournot model seems more appropriate. The exchange functions as an auction that clears the market given the quantities that are bid in. Furthermore, the Bertrand equilibrium is supported by the assumption that any firm can capture the entire market by pricing below others, and can expand output to meet such demand. Since generation capacities present significant constraints in electricity markets, this assumption is not tenable. Previous research suggests that if firms choose their capacities and then compete on price, within the restrictions of their capacity constraints, the outcome is most closely approximated by the Cournot model.[12]

Thus, the centralized price mechanism and capacity-constrained suppliers in electricity markets (at least during peak periods) support the use of a Cournot model for a base-case analysis. As we will see in the next two subsections, it probably overstates the amount of market power, but such a conservative approach seems defensible for an

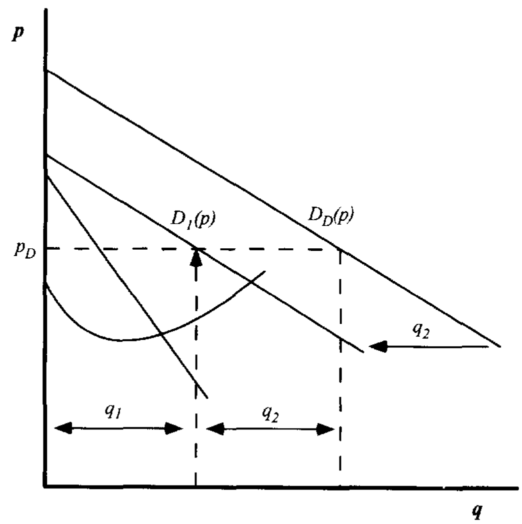

Figure 2. Symmetric Cournot Competition

initial analysis. The outcome of such an analysis would be an estimate of the Lerner index of market power, which measures the markup over the perfectly competitive price. Such conclusions would depend upon the price elasticity of demand, the capacities and cost curves of Cournot competitors, and the cost curve of any competitive fringe.

We will illustrate this analysis with a simple model. Two duopolist firms face a downward sloping demand curve, $D_D(p)$, that is composed of actual demand minus the amount supplied by the competitive fringe. Figure 2 illustrates the strategy of one of two identical Cournot firms in response to the behavior of its competitor. Firm 1 conjectures a supply quantity that will be provided by its opponent. The opponent's supply quantity, q_2, shifts the 'residual' demand curve that firm 1 faces inward from $D_D(p)$ to $D_1(p)$. Just below, its associated marginal revenue curve, MR_1, is shown intersecting firm 1's marginal cost curve at the profit maximizing output level. In equilibrium, q_1 will be the firm's optimal response to q_2 and vice versa.

Mathematically, we can write the profit function, π_i^c, of any Cournot competitor as a function of its own output and its competitor's output levels, and the inverse demand function, $p(q)$, derived from $D_D(p)$.

$$\pi_i = q_i p(q_i + q_0) - c_i(q_i) \tag{2}$$

where q_i and $c_i(q_i)$ are, respectively, the output and cost of firm i, and q_0 is the output of all other competitors. A first-order condition for maximizing profit, given q_0, is therefore

$$p(q_i + q_0) + q_i p'(q_i + q_0) - c_i'(q_i) = 0 \tag{3}$$

Simultaneously solving Equation (3) for all firms will give the Cournot equilibrium. This would provide a first-order approximation of the severity of market power resulting from the concentration of generation sources. The analysis would require estimates of the cost curves of the oligopoly competitors, the costs and capacity of the fringe, and estimates of demand elasticity over the appropriate time frame. As will be seen in the next two sections, two important considerations indicate that the Cournot solution can be thought of as a 'worst-case' possibility for non-collusive outcomes.

Supply curve competition

Cournot competition does not fully describe the options available to the firms in the electricity market. Generators are not forced to bid quantities in a spot market, but are in fact free to bid any supply curve, with a quantity bid corresponding to the special case of a vertical supply curve. However, an estimate of a static Cournot equilibrium of the electricity market would still provide a rough estimate of competitive behavior if firms face little demand uncertainty. When there is no uncertainty, it turns out that of the many Nash equilibria that are possible, the one produced by quantity bids (the Cournot strategy) is the most profitable. Thus, if there were no uncertainty, and cost data were available, an estimate of the price–cost margin from the Cournot model could take the place of a structural index such as an HHI calculation.

In a market with uncertain demand, the situation is different. A producer will face many possible demand levels, even when it knows its competitor's production levels. Firms then engage in supply curve competition. This problem was analyzed by Klemperer and Meyer[13] for a general context. Under supply curve competition, it is profitable for firms to move away from the Cournot equilibrium towards a Nash equilibrium that is described in terms of upward-sloping (and definitely not vertical) supply curves. Thus suppliers are not bidding simple quantities as specified by a Cournot model. This outcome leads to price–cost margins that are smaller than those from Cournot competition. The introduction of demand uncertainty therefore mitigates the effects of market power.

Green and Newbery[14] adapted supply curve competition to the context of the electricity spot market in the UK. In this case, the source of demand variation was the bidding process itself. Each generator bids a single supply curve to cover an entire day. As demand fluctuates, hourly prices are set by moving up and down the generators' supply curves. From the bidder's perspective, the outcome is the same as in the Klemperer and Meyer framework: a single curve will determine various outcomes subject to random fluctuation. Bolle[15] also adapted supply curve competition to electricity spot markets, with particular emphasis on collusive outcomes. Since such competition best represents the strategic options available to firms in the spot market, the supply curve analysis, while more complicated, would be more informative than a Cournot model.

Entry and market power mitigation

The simplified example in the second section assumed that demand is inelastic and did not account for the possibility of a new entry of suppliers. While we have argued that such assumptions are somewhat justifiable in the context of an isolated, 1 hour market for electrical energy, they are clearly not appropriate for a more general analysis. Even a monopolist with no competitive fringe who exploits the short-term inelasticity of its demand will eventually see customers turning to, or developing substitutes for, that product. In estimating a Cournot or any other oligopolistic equilibrium, longer-run elasticities of demand and fringe supply should also be analyzed.

The threat of competitive entry affects a dominant firm's behavior in much the same way as does the existence of a competitive fringe. A rational dominant firm would be cognizant of the fact that the higher it raises prices, the more entry into that market it will stimulate. The firm would then account for this threat of entry by observing that it faces a residual demand curve, where higher prices stimulate more supply from new sources. If the dominant firm has a cost advantage over potential entrants, it may be optimal for that firm to set prices just below the level where entry would occur. The threat of entry can therefore have an effect on the pricing of even a purely monopolistic firm. A practical analysis of the prospects for entry should include the manner in which some costs of existing facilities are recovered through a transition charge, and whether such cost recovery gives the incumbent firms a cost advantage. Entry may also depend on the availability of risk-reducing financial instruments, a prospect we discuss shortly.

Collusion and the bidding process

Up to this point, we have discussed only non-cooperative models of competitive behavior, assuming no explicit coordination between competitors. However, the possibilities for, and implications of, collusion between parties should also be examined. While Joseph and Marcus[16] have observed that most forms of explicitly collusive behavior in these market would be illegal, the ability of regulators to detect and punish such behavior in a timely manner cannot be assumed to be unlimited. Collusion has been found in many markets where it is illegal.

It is possible that some unusual characteristics of the electricity spot market may lend themselves to collusive practices. The strategic actions of most firms will be fairly transparent, although many aspects of the bid process pertinent to the question of collusion, such as what bid information is made publicly available, have yet to be established. Spot market auctions will be repeated frequently, and this could provide opportunities to 'punish' firms who deviate from collusive equilibria.

Transmission congestion and the market for peak power

The ability to exercise market power depends largely on the number of competitors and their shares of the market, and this in turn depends on the geographical extent of the market. However, past trading patterns from an era of rate-of-return regulation may not be a good guide to the trades that are likely to be economic in a restructured market. Given a California pool and open-access tariffs mandated by the Federal Energy Regulatory Commission,[17] not only will historic trading patterns become less relevant, but all institutional barriers to trade will become much less important. Consequently, the relevant definition of geographic markets will come to be based primarily on physical characteristics of the transmission grid.

Two physical characteristics of the grid are of primary importance when defining the market's geographical extent: line losses and congestion. These two characteristics are quite different in their effect. Congestion tends to operate locally, while line losses increase with distance. Congestion is a fairly discontinuous phenomenon, while loss is quadratic in power flow. In one way, however, they are quite similar: both become more restrictive at higher power flows. Consequently, both cause the market size to shrink at high usage levels, and to expand at low usage levels. Because we think line losses present less of a problem, this section deals only with congestion and its effects. We make the simplifying assumption that the network is lossless.

Congestion and geographic market boundaries

Setting aside line losses and institutional constraints, the geographical extent of the market is limited only by congestion. Without congestion, the physical market could then include all of the USA and

Canada. In fact, when the demand is very low, the entire western system can be essentially free of congestion and could, under open access, form a single market. Such a market can be treated as if it consisted of a single node, which would severely limit the market power of any trader. Since this is the uncongested state of the market, we can see that generation market power is almost entirely dependent on transmission congestion. Congestion serves to fragment the physical market, thereby creating the possibility of market power in subregions.

Congestion fluctuates continuously, so market power does as well. This significantly raises the level of complexity of a market-power analysis. Joskow's tests[18] for 'distinguishing firms and markets that are at "low risk" for excessive market power' must be reformulated. At the most off-peak hour, every market will pass this test, but at the times of highest demand, many markets will fail it. Yet, if the market fails the test only 1 hour per year, should that raise policy or antitrust concerns? Probably not. This leads to the conclusion that any static test of market power cannot be considered well specified until it is associated with some annual duration. Duration is a particularly difficult part of the specification because, as has been demonstrated in the UK, under some circumstances firms can earn substantial profits in a very short period of time.

The two cases of congestion affecting market power

Although there is no limit to the complexity of geographical separations that can be caused by weak lines and congestion, two cases illustrate the basic issues that can arise:

(1) flow on the congested line simply acts as a shift in demand at each end of the line, and
(2) suppliers at each end strategically respond to threats of competition from suppliers at the other end.

We consider each in turn, and represent each by the simplest possible example. Each example has only two nodes and a single connecting line that is congested or is in danger of being congested. The first example has market power at only one end, while the second has market power at both. There are, however, more complex examples of case 1 which do have market power at both ends, so it is not as easy as it might seem to tell which case applies in any given market situation.

In the California setting, the two cases may correspond to congested lines that enter the state from the north and east, connecting it to large external markets that are essentially competitive (case 1), and to congested lines inside the state and connecting regions that have a limited number of suppliers, and thus exhibit market power at both ends of the lines (case 2). Both cases are important in understanding market power in California and the potential for reducing it.

Case 1: Two-node monopoly-competition

This case illustrates the most basic point about congestion and market power. The inflow of power on the congested line is equivalent to demand shifts at both ends of the line. Firms at the transmitting end face increased demand, and firms at the receiving end face reduced demand. If, as in our example, there is a monopolist at the receiving end, it may mark up prices less, but it will still have some degree of market power.

In this example we call the two nodes East and West, with the East node assumed to be competitive and of sufficient capacity to congest the line to the West. The West node represents California, and is assumed to be subject to market power on the part of suppliers. To simplify analysis we represent this market power as a monopolistic supplier. We also assume for simplicity that both the monopolist and the competitive suppliers to the East have the same constant marginal cost.

Because the East node is competitive, price there will equal marginal cost. The monopolist in the West, however, will always find it profitable to restrict demand until price exceeds marginal cost. Consequently, with price higher in the West, profit-maximizing behavior of the competitive firms in the East will cause the line to be congested with power flow from East to West. Once the line is congested, the monopolist knows that there can be no response from the East to a marginal change in the quantity that the monopolist supplies. Thus, the monopolist views its demand curve to be $P(Q + k)$, where $P(\cdot)$ is local demand and k is the fixed congested flow from the East. The single firm in the West is simply a monopolist facing this residual demand. Because demand is lower, the monopolist will set a lower price, so in this sense its market power will be reduced. Still, on the margin, the monopolist does not take into account any competitive behavior of firms to the East.

The existence of the line causes the Eastern node to face increased demand. In this example, the demand increase does not affect price because the node is competitive and marginal cost is constant. If the industry supply curve in the East were upward sloping, then the line could increase prices in the East. Similarly, if there were market power in the East, this demand on the congested line might cause a price increase relative to the price with the line absent. Again, the Eastern-node firms would not need to account for any competitive behavior on the part of the Western monopolist. The fully congested line effectively insulates each node from the other's strategic behavior.

Case 2: Two symmetric nodes and suppliers

The simplest example of case 2 resembles one of the main constraints that will exist in a California power pool. This is the constraint that occurs when transmitting power from Northern to Southern California (the so-called 'south of Tesla constraint'). This north–south constraint probably determines the two principle California regions relevant for market power analysis. In this section we make the simplifying assumption that the two regions are identical in every respect, and that there are monopoly suppliers in each.

If these two regions were completely separate, there would be identical monopolistic solutions in the two regions. If we connected them with a very strong line, so that each supplier could sell as much as it wanted in the other's market, there likely would be a symmetric duopoly solution with lower prices. Although this result is obvious, it has one very surprising property: no power flows on the connecting line. This is a consequence of the complete symmetry of the problem. This means that although the line is not used, it is still very useful, because it keeps prices low. The threat of competition is all that is really needed, and the line (if it is big enough) provides that threat. So, we conclude that if a connecting line is of sufficient capacity to reduce market power as much as possible, it may appear to be overbuilt and underused.

This raises the question of how big a line is needed to induce duopolistic, instead of monopolistic, behavior. Clearly, the answer need not be related to the actual power flow on the line. More generally, we would like to know the effect of any given size line on the degree of competition between the firms at each end of the line.

If we imagine the two nodes connected by an extremely weak line, then each supplier would almost ignore the other, because each would know that the other could affect it very little. Thus, we would have something close to the two-monopoly solution in spite of the connecting line. The actual outcome, however, is more complicated than this, because it is not optimizing behavior for either firm to leave the line unused. The research we have done so far indicates that there will, in fact, be no pure-strategy equilibrium for monopoly markets connected by a very thin line. The Cournot equilibrium would involve each firm randomizing between an aggressive strategy – in which it attempts to export to the other market – and a passive strategy – in which it allows the other firm to send power into its market up to the capacity of the line and then maximizes profits on the resulting residual demand.

As the line capacity increases, it appears that the probability-weighted average price – averaged over the possible outcomes from the mixed strategies – declines. If the line is thick enough, then the result will be the pure-strategy, unconstrained Cournot equilibrium. The size of the line necessary for such an outcome may be surprisingly small relative to the increase in output it stimulates. In the example that we have worked out[19], the capacity of the line necessary to increase output in each market from the monopoly to the Cournot level is only about one-sixth of the total increase in output. Again, in this symmetric example, no power actually flows on the line.

The market for contracts

The potential for abuses of and solutions to market power in the spot markets could be greatly influenced by the competitive condition of a 'contract market' in electricity. Technically, the term 'contract' would apply to any transactions made outside of the spot market: that is to say, any arrangement for which the agreement to transact is reached at a time different from that in which actual delivery occurs. These transactions may involve the future delivery of electricity or financial compensation based upon market outcomes. We used the phrase 'contract market' to describe the collection of agreements that could be made around, or with reference to, the exchange-based spot market. In this section we discuss three of the most prominent such instruments: physical power contracts (PPCs), contracts for differences (CFDs), and transmission congestion contracts (TCCs).

Contracts for differences. CFDs are basically insurance or hedge contracts. Two parties agree on a price for electricity and a quantity to be 'covered.' If the spot price turns out to be greater than stated in the contract, then one party pays the other the price difference for each unit of the covered quantity. If the spot price turns out to be lower than stated in the contract, the opposite party pays the difference. Thus, these contracts would be treated by most parties as the equivalent of forward or futures contracts that are available for many commodities, including crude oil and natural gas.

The market for financial contracts has, at first glance, a vast pool of potential participants and few

Table 2.

Payment (for 1 kwh)	Agent: Power consumer	Power generator	'Long' speculator
Spot market (with price p)	$-p$	p	–
CFD: Generator pays customer difference between p and p_c	$-p_c + p$	$p_c - p$	$-p_c + p$
Net payment	$-p_c$	p_c	$-p_c + p$

barriers to entry. Since no physical product is involved, third parties, such as banks and insurance companies, could in theory enter this market with little difficulty. However, the outlook for competition in the contract market is not as clear cut as at first glance. Firms involved in the physical power market may have a natural interest, which could even be thought of as a negative cost, in participating in the market. Buyers in the power market, for instance, are likely to be the lowest-cost participants in CFDs on the long side of the market, ie the party that pays out if the spot price is lower than expected and receives payment if the spot price is higher than expected. Such a position would operate as a simple hedge of the buyer's purchase price risk (see Table 2). Unlike third parties, which would have to be paid an expected premium to be on the long side of a CFD (through a contract price, p_c, that is below the expected spot price), risk-averse power buyers would benefit from the position even if the contract price were somewhat above the expected spot price. In other words, the power buyer may prefer to pay p_c with certainty than an uncertain price with an expected value less than p_c.

In this way, buyers of power can take 'long' positions in the CFD market at lower cost than speculators. The likelihood that power suppliers, who would like to be on the short side of CFDs to hedge their risk, can do so without paying a premium therefore depends upon the degree of competition among the low-cost, long-side participants in the CFD market, ie the power buyers.

Financial institutions or other third parties are at some disadvantage because their participation in the contract market would likely be risk increasing rather than risk decreasing (hedging). Besides bearing the 'normal' risk from unforeseen supply and demand conditions, speculators may be more exposed if manipulation of the spot market were to occur. Hedgers, who deal in the real, not just the financial product are naturally insulated to some extent from extreme price behavior. If manipulation were a frequent or serious problem, this could drive the speculative institutions out of the contract market.

The terms available to short-side participants in the CFD market are likely to be critical for fostering competition in power supply. The price risk of building a new power generator could be greatly mitigated in a CFD market, and the ability to do so at low cost could very well determine whether a new entrant was likely to be profitable. Two factors will affect the terms of the CFDs available to short-side participants in the CFD market: (1) the cost advantage of power buyers versus third parties in being on the long side of these contracts, which will depend in part on the proportion of the risk that is diversifiable and the cost of doing so; and (2) the concentration of power buyers. Essentially, how large will the power buyer's cost advantage be in this market, and how much of the cost advantage will be passed along to short-side participants?

Transmission congestion contracts. A special form of exchange-based financial instrument that may require individual attention as a product market is the transmission congestion contract (TCC). In its December 1995 decision, the CPUC indicated that TCCs would form the basis for the allocation of transmission congestion rents amongst the owners of transmission assets. A TCC, which is characterized by a quantity level and two nodal locations, entitles its owner to receive a revenue stream equivalent to the price difference between those two locations, times the specified quantity. Thus a TCC of quantity t from nodes 1 to 2 of a network would pay its owner $t(p_2 - p_1)$. (See Bushnell and Stoft[20] for a more detailed analysis of CFDs and TCCs.) TCCs can be used by traders to hedge locational price risk in the same way that CFDs are used to hedge price fluctuations over time.

Physical power contracts The CPUC's December 1995 decision allows contracts for the 'physical' delivery of electricity to be implemented in combination with an exchange-based market. The pool of potential providers and consumers of this product is much smaller than that for financial-based instruments. Since the product involves the actual production/acquisition or consumption of electricity, only those in a position to provide or consume this energy can enter into an agreement. From this perspective the potential suppliers of this product are exactly the same as for the spot electricity market.[21] However, the discussion in this section of the influence of contracts on the exercise of spot market power also applies to markets that include physical contracts. The topics we address are the entry of new generation, and the 'temptation' of an oligopolist to sell forward some of its production. For our purposes, it does not matter whether this is accomplished through physical or financial forward arrangements.

Interactions between spot and contract markets

There has been much recent attention paid to the competitive implications of interactions between long-term financial markets and their underlying commodity spot markets. The primary hypothesis of this body of work is that a competitive contract market can undermine the ability of dominant firms to exert market power in the spot market. Such arguments are based upon either the game-theoretic interactions among existing suppliers[22] or the use of contracts to facilitate new entry into a concentrated market.[23]

As Allez and Vila[24] demonstrated for a stylized duopoly model, the existence of a viable forward market can create a 'prisoner's dilemma' of sorts for oligopoly suppliers. Each player could make additional profit by moving first into the forward market and stealing away some demand from the other suppliers. However, when all suppliers act in this way, the total quantity supplied increases towards the competitive levels, undermining the oligopoly equilibrium that existed before the forward market came into existence.

A second line of economic theory, however, indicates that this beneficial influence of contract markets on spot markets is overstated or even reversed. In modeling oligopoly equilibria for durable goods, Gul[25] and Ausubel and Deneckere[26], show not only that a Cournot duopoly outcome can be sustained, but also that the monopoly outcome could be achieved. The key to their results is the multiple periods in which the product could be sold. In essence, the multiple periods produce a repetition of the 'prisoner's dilemma' described by Allez and Vila.[27] In a repeated prisoner's dilemma, one competitor can 'punish' the other for deviating from the cooperative outcome. The threat of punishment induces each competitor to cooperate with the other and sell quantities that add up to the monopoly output level. While electricity is certainly not a durable good, the markets we are describing will be repeated every hour of the year. This repetition may allow competitors to cooperate implicitly to reduce output.

While the effect of contract markets on existing producers is in dispute, evidence presented by Newbery[28] and by Wolfram[29] on the UK electricity industry seems to bear out the hypothesis that they play an important role in facilitating the entry of new producers. In the UK, the perceived threat that the two dominant suppliers would artificially inflate pool prices indefinitely led regional electric companies to sign contracts for differences with independent power producers at a price somewhat above the competitive spot price but below the expected oligopoly price. As Newbery points out, the rational response of existing suppliers would be to enter into contracts themselves at a price just below that which would cause entry. The fact that the generating companies did not do this could indicate a weakness in this hypothesis or could simply reflect myopic behavior on the part of the Gencos. There may also be a self-dealing element in the RECs contracts, as they were allowed to hold equity states in the generation companies that they were contracting with. Regardless of this fact, some of the theoretical and empirical evidence now indicates that thriving contract markets can mitigate market power abuses in spot markets. However, many questions regarding the specifics and degree to which this effect would arise in California electricity markets still need to be resolved.

Electricity contract markets in California

The key concern regarding contract markets in California is the concentration of market power in the hands of large distribution companies (Distcos) that are also affiliated with generation companies. This concentration alone raises the concerns discussed above that a large buyer, or a concentrated group of buyers, will be unwilling to enter the long side of a CFD without expected compensation that considerably exceeds its cost of entering the contract. Under current proposals, these regulated Distcos could use this method to deter the entry of competition for their generation affiliates.

In the UK, it was the distribution companies that, by entering into contracts with independent generators, facilitated entry that is undermining the market power of the generating companies. A long-term contract for difference allows independent power producers to reduce risk and lower the cost of capital, which otherwise might not have been available. A Distco with affiliated generation is unlikely to want to undermine the market share of its own affiliate or facilitate the entry of potential competition for that affiliate. Affiliation may also affect the oligopolistic interaction between suppliers and demanders to the detriment of competitive efficiency. If it is true that a robust forward market helps undermine oligopoly in the spot market, Distcos and their generation affiliates may wish to prevent a viable forward market from forming. Those issues are significantly more complex and require further examination.

Affiliations between Distcos and Gencos could cause similar problems in the market for transmission congestion contracts. Initial proposals have suggested that TCCs will be awarded to current owners of transmission facilities to compensate them for their investment in these facilities and insulate them from future congestion on these lines. The Distcos created out of the current IOUs would therefore initially own the lion's share of these TCCs. If TCCs become an important hedging instrument, acquisition of a TCC could lower the cost of entering the generation market. Then the Distcos would have an incentive to protect their affiliates by withholding TCCs from new generation projects. As with CFDs, other firms could offer to hedge locational price risks, but would probably command a steeper risk premium and prove more costly than would TCCs.

Currently, it is unclear how the presence of large customers using CFDs or signing physical contracts will affect this competitive outlook. Large customers, and eventually all customers, will have the right to directly access the spot market or sign physical delivery contracts. To determine the extent to which this 'bypass' of the Distcos might compensate for a failure in the financial contract market, it would be necessary to examine the potential demand levels of those customers that are allowed to enter into such agreements, and the transaction costs involved with pursuing direct supply arrangements. These customers may form only a fringe or could eventually create robust competition for contracts.

The market for reliability services

In addition to the markets that will be directly associated with the exchange-based provision of spot electricity, we expect new markets for products traditionally bundled with 'electric service' to emerge. Prominent among these products are those required for ensuring the integrity of the network and the stability of the grid. These include markets for load balancing, spinning reserve, and voltage support. While it is clear from the various proposals and decisions currently on record that these products will be necessary for the operation of any form of competitively restructured electricity industry, little detail is known at this time about how these markets will be developed and integrated with the basic market for spot power. Since any representation of these markets would be far more speculative than even our models of the spot power and contract markets, our examination of market power concerns in these markets is more general, and is related to the nature of the product rather than the specific mechanics of the markets.

In this section, we cover some broad issues associated with the implementation of these markets that we think will be important in determining their competitive outlook. The most important distinction between products in these markets for 'reliability services' is that between markets whose competitive outlook closely resembles that for spot power and those markets that may have suppliers with significant geographic market power. The markets for spinning reserve and load balancing seem to share many key characteristics with the markets for spot power, while the market for voltage support has unique geographic requirements that create significant potential for geographic market power.

Reliability services with the same geographic markets as spot power

The exchange-based spot market will apparently require supply-curve bids from generators 24–36 hours in advance. There will inevitably be deviations from these supply curves due to forced outages, weather conditions, or even intentionally inaccurate bids. There are two immediately obvious possible solutions to this problem of imbalances. One is to readjust the spot price to a level where actual supply meets demand, thereby possibly calling on suppliers not originally scheduled to generate. This approach has both the attraction and the difficulty of requiring establishment of mechanisms for consumers to respond to price changes by stating in advance the conditions, price or other, under which they agree to be curtailed. The other solution is to create a separate market whose participants would compete specifically to meet these shortfalls on little notice.[30] A similar but potentially distinct market may be created to provide moment-to-moment load balancing during the half-hour or hourly periods in which the spot price is fixed.

In addition to short-term load balancing, system operators will presumably require a level of spinning reserve to ensure the continuous stability of the grid. This product could be integrated with the provision of short-term load balancing if that market was thick enough. Spinning reserve is, after all, a tool that is used to guarantee that loads can be balanced in the face of a major contingency. If the market for instantaneous energy is deemed unreliable, however, a separate 'spinning reserve' market would be needed. Again, the potential for demand-side participation is important, and requires further development.

The competitive outlook in markets for load balancing and spinning reserve appears to be very similar to the outlook for the spot-power markets. Peaking generation technologies with short downtimes and high ramping rates may have a cost advantage in these markets, so an analysis of the potential suppliers of this type might be useful. However, these products could also be provided by baseload or intermediate generation that cycles below its maximum output level. Any artificial rise in the price levels for these services would therefore spur entry from suppliers to the spot market. As with the spot market, transmission congestion could isolate certain geographic markets. Spinning reserve is of little use if it is on the wrong end of a congested line.

One case that should be of concern is a transmission path that is not congested but does not have enough capacity to provide sufficient additional service to meet a contingency. This scenario is discussed in the following section.

Reliability services with potential for local market power

As Joseph and Marcus (Ref. 15, p. 21) describe,[31] there are currently about a dozen oil/gas generation units in the state that are considered 'must run' units for reliability purposes. These include units in the San Francisco and Humboldt Bay areas that are operated in this fashion due to transmission constraints, and units on the Central Coast, East Bay and in San Diego that are used for voltage support. As described above, a region whose transmission lines are not congested may still lack the transmission capacity to meet all demand in that region in the event of an outage. Under these conditions, local generation has market power because little effective substitution from competition is possible.

The provision of voltage support is, by definition, a geographically specific product. It is voltage at specific locations (most frequently demand centers) that must be maintained. It would be difficult to organize competition for voltage support if it were to be provided exclusively in the form of reactive power produced by generators. The reason for this is that reactive power is a joint product with real power, and it is difficult to 'transport' reactive power over the network. This tends to confer market power on those generators suitably located to provide this service. A pricing approach to reactive power has been proposed.[32] But the local nature of the market implies that pricing can produce distortions and inefficiencies.[33] Current utility practice in California addresses the voltage support requirements by dispatching local generators (at least up to their minimum operating levels) out of merit order, ie on a 'must-run' basis. Under cost-of-service regulation, these generators are paid their operating costs. In a competitive setting, there are potential alternative approaches to this problem focusing on the demand-side, which we discuss in the next section. Finally, it is also possible that some geographic markets for spinning reserve will be smaller than their corresponding on-peak power markets.

Demand-side and other alternatives for reliability services

The greatest potential for new entry into the reliability services markets may come from demand-side alternatives. Such developments could allow for significant improvements in the efficient operation of an electric system. This potential benefit of restructuring has not drawn nearly as much attention as it deserves. Interruptible demand could be used to account for imbalances between the day-ahead and current supply and, once the proper metering and communications devices are in place, even account for within-hour variations. Widespread implementation of a market for loads that are price sensitive even in the very short run can lead to adjustments or even redefinitions in the requirements for spinning reserve, which is currently utilized to ensure that no customers lose service.

An alternative approach to local voltage support problems is the widespread adoption of capacitors, or more sophisticated reactive compensation devices at load centers. These devices would eliminate, or substantially reduce, the need for reactive power produced by generating units. This strategy has been adopted by Virginia Power.[34] Under this approach, generators still provide reactive power in system emergencies, but their ordinary dispatch is not constrained by steady-state voltage support requirements.

The major market-power issue with respect to these new sources of reliability service products is the potential for distribution companies to discourage the development of these products. Distribution companies will be the dominant purchasers of energy, at least in the early phases of the implementation of these markets. Any widespread adoption of flexible, price-sensitive demand will depend upon the willing and active participation of Distcos in developing such programs. As with the contract market, the affiliation between suppliers and consumers here could lead to an incentive to suppress, rather than encourage, such developments.

In the case of a demand-side approach to voltage support, a competitive entry mechanism would not suffice. This approach is predominantly a grid standard, and would have to be adopted as part of the overall market rules. If such an approach were adopted, the question of allocating the costs of reactive compensation devices would need to be settled. Additionally, there would need to be a mechanism to set technical standards that determine exactly what demand-side measures would be appropriate. If emergency services would still be required of generators, it would be necessary to determine whether these services would be procured under some kind of contract, or whether they would simply be required and subject to *ex post* cost compensation.

Conclusions

We have examined the markets that are likely to be the most significant in a restructured electricity industry. In doing so, we have highlighted some of the potential market power concerns in these markets and outlined methodologies for examining the potential severity of these concerns. The key to assessing market power will be a correct definition of the geographic extent of these markets. Geographic distinctions should be made on the basis of the potential for transmission congestion, which could isolate various regions from outside competition. Obviously, the level of congestion will depend on the levels of demand. In market power analysis, distinctions will therefore have to be made between high- and low-demand periods. Pool-based financial instruments – such as contracts for differences and transmission congestion contracts – and reliability services – such as load balancing, voltage support, and spinning reserve – are other products that will play important roles in the future electricity industry.

Structural measures of market power, such as the HHI, traditionally have been employed for initial analyses of markets. Such measures have certain general shortcomings, some of which are exacerbated when applied to restructured electricity markets. Fortunately, much more cost and demand data are available for these markets than most others. Thus, while structural indices appear to be less useful here, directly calculating estimates of competitive equilibria, such as a Cournot oligopoly equilibrium, appears to be more feasible for this industry than is generally the case. The calculation of the possible equilibria in these forthcoming markets can provide important, although rough, indicators of the potential for inefficient outcomes in these markets.

While reliability of supply will continue to be a goal and a concern in the restructured market, explicit payments for *capacity* will probably not be the best way to achieve that reliability. The evidence from the UK, which has incorporated a capacity payment into its pool market, indicates that the mechanism designed for that setting is vulnerable to manipulation. Instead, competitive markets to provide reserve supply or allow interruptible demand could provide the desired level of reliability. More generally, capacity payments do not, in general, provide the appropriate signals for efficient investment. The economic incentives for investment can be provided by spot markets for energy and reliability, and associated financial instruments.

Despite the large degree of uncertainty that remains about the impacts of market power in the emerging electricity industry, our analysis illustrates several areas that will have a key impact on the competitive health of this industry. Beyond assessing the immediate prospects for market power, policy makers should therefore also focus on the best prospects for solving the market power problems that do emerge.

One such area of focus is the price sensitivity of demand. It is widely perceived that short-run demand for electricity is very inelastic. There are indications that technological and regulatory barriers have played a significant role in making it so. It seems clear that significant problems may result if the industry is deregulated without any improvement in the price-responsiveness of demand. We have shown that flexible demand can play a central role in mitigating the effects of market power and in driving efficient investment. Policy makers should therefore take a strong interest in improving short-run price responsiveness. The concentration of demand within large distribution companies, however, may create barriers to demand side innovation if those distribution companies are also affiliated with generation companies.

Unfortunately, the impact of restructuring on the behavior of consumers is one of the areas where a large amount of uncertainty remains. Up till now, regulatory and institutional moves towards competition have been largely driven by technology advances. Cheaper and more dispersed generation technologies have placed enormous pressure on large utilities both to restructure internally and to open themselves to competition. Demand-side technologies, however, remain more in the experimental stage. Thus for the first time in the restructuring process, regulatory initiatives will precede the installation of the technology required to make those initiatives succeed.

Another area that should be of concern to policy makers is the emergence of competitive markets for contracts, either for the physical delivery of power or for exchange-based financial instruments. The availability of such contracts will directly affect the ability of new generation suppliers to enter the spot and reliability markets and therefore to alter the level of competition for those markets. Conversely, the consolidation of demand and/or transmission congestion contracts in the hands of distribution companies with generation affiliates could artificially raise the price of such contracts and constitute a barrier to entry into the spot markets.

Beyond the distribution of TCCs, transmission networks in general constitute a third area on which policy makers would focus. The physical configuration of transmission networks will impact the nature of competition in electricity markets in ways that are currently not well understood. The patterns of transmission congestion experienced in today's regulated markets may not always give a good indication of the nature of congestion once competition becomes more widespread. Beyond simply allowing cheaper power to flow into a market, transmission lines can also play a strategic role by making threats of competition credible. Thus even a line that experiences little net flow can play a very important role in mitigating market power.

The authors would like to thank Carl Blumstein, Richard Green and Bart McGuire for helpful comments and discussions. This research was partially funded by the California Energy Commission under Contract No. 700-93-003, amendment 2.

[1]There are, however, a number of foreign precedents from which we can draw experience, namely Norway, Latin America, and the UK.

[2]Paul L Joskow, 'Horizontal market power in wholesale power markets.' Comments in FERC docket no. RM95-9-000, 1995.

[3]Adam Jaffe and Frank Felder, 'Should electricity markets have a capacity requirement: If so, how should it be priced?' Mimeo, Economics Resource Group, Inc. Cambridge, MA, 1996.

[4]D Levy and E Kahn 'Accuracy of the Edgeworth approximation for LOLP calculations in small power systems', *IEEE Transactions on Power Apparatus and Systems.* Vol PAS-101, No. 4, 1982, pp 986–996.

[5]David M Newbery, 'Power markets and market power', *Energy Journal,* Vol 16, No 3, 1995.

[6]In the future, large generation companies in the UK will be required to report their availability policies for the upcoming year. It is unclear at this time whether this will be an effective deterrent to the strategic games.

[7]For instance, home security companies charge for the 'on call' service. In some cases, they also charge a 'per visit' fee when the contingent demand for their service arises.

[8]California Public Utilities Commission, Decision 95-12-063 as modified by D.96-01-009, 20 December, 1995.

[9]James Bushnell and Shmuel S, Orens 'Bidder cost revelation in electric power auctions', *Journal of Regulatory Economics*, Vol 6 No 1, 1994, pp 5–26.

[10]While the vertical integration and transmission constraints may create concerns that incumbent firms will be able to discourage new entry, it seems unlikely that large firms could engage in profitable predation to drive out firms that are already operating. Profitable predation requires that the predator increase output substantially to drive price below the prey's short-run average cost, but then reduce output once the prey is eliminated. Because short-run average cost is generally quite low once an electricity-generating facility is operating, this would require a significant depressing of price for a significant period of time. Furthermore, the predator would need to have sufficient capacity to expand output, capacity that it would not want to use (in this market) after the predatory behavior ends.

[11]Jean Tirole, *The Theory of Industrial Organization,* MIT Press, Cambridge, MA, 1988.

[12]David M Kreps and Jose A Scheinkman, 'Quantity precommitment and Bertrand competition yield Cournot outcomes', *RAND Journal of Economics,* Vol 14, No 2, 1983, pp 326–337.

[13]PD Klemperer and MA Meyer, 'Supply function equilibria in oligopoly under uncertainty', *Econometrica*, Vol 57, No 6, 1989, pp 1243–1277.

[14]Richard J Green and David M Newbery 'Competition in the British electricity spot market', *Journal of Political Economy*, Vol 100, No 5, 1992, pp 929–953.

[15]Friedel Bolle, 'Supply function equilibria and the danger of tacit collusion', *Energy Economics,* Vol 12, No 2, 1992, pp 94–102.

[16]Marc D Joseph and David Marcus, 'Comments of the Coalition of California Utility Employees' in *CPUC* OIR.94-04-031 and OII.94-04-032, 23 July, 1995.

[17]California Energy Commission, '*Generation Market Power in Electricity Restructuring,*' Sacramento, CA, 1995.

[18]Joskow, *op cit*, ref 2.

[19]The example has identical linear demands in the two markets and zero marginal cost.

[20]James Bushnell and Steven Stoft, *Transmission and Generation Investment in a Competitive Electric Power Industry,* PWP-030, Program on Workable Energy Regulation, University of California Energy Institute, Berkeley CA, 1995.

[21]It is important to note that for the first 5 years after restructuring, the three California IOUs are required to supply their energy only through the spot market, and are therefore not eligible suppliers of PPCs. The concentration of suppliers for this transition period may be drastically different, although spot electricity combined with financial instruments remains available as a substitute for PPCs.

[22]Blaise Allez and Jean-Luc Vila, 'Cournot Competition, forward markets and efficiency', *Journal of Economic Theory*, Vol 59, No 1, 1993, pp 1–17; Richard J Green, 'Contracts and the Pool: The British Electricity Spot Market.' mimeo, Department of Applied Economics, Cambridge University, 1992.

[23]Newbery, *op cit*, ref 5.

[24]Allez and Vila, *op cit*, ref 22.

[25]Frank Gul, 'No cooperative collusion in durable goods oligopoly', *RAND Journal of Economics,* Vol 18, 1987, No 2, 1987, pp 248–254.

[26]Lawrence Ausubel and Raymond Deneckere, 'One is almost enough for monopoly', *RAND Journal and Economics*, Vol 18, No 2, 1987, pp 255–274.

[27]Allez and Vila, *op cit*, ref 22.

[28]Newbery, *op cit*, ref 5.

[29]Catherine D Wolfram, 'Measuring duopoly power in the British electricity spot market', Mimeo, Dept of Economics, MIT 1995.

[30]Another potentially contentious question is what penalties should be paid by suppliers who deviate from their bids. We do not address this question here.

[31]Joseph and Marcus, *op cit*, ref 15, p 21.

[32]William Hogan, 'Markets in real electric networks require reactive pricing', *Energy Journal*, Vol 14, No 3, 1993, pp 171–200.

[33]Edward Kahn and Ross Baldick, 'Reactive power is a cheap constraint', *The Energy Journal*, Vol 15, No 4, 1994, pp 191–201.

[34]P Nedwick, A Mistr and E Croasdale, 'Reactive management: a key to survival in the 1990s', *IEEE Transactions on Power Systems,* Vol 10, No 2, 1995, pp 1036–1043.